

Common information revisited

Ilya Razenshteyn*

Moscow State Lomonosov University, Mathematics Department,
Logic and Algorithms Theory Division, ilyaraz@gmail.com

Abstract. One of the main notions of information theory is the notion of mutual information in two messages (two random variables in Shannon information theory or two binary strings in algorithmic information theory). The mutual information in x and y measures how much the transmission of x can be simplified if both the sender and the recipient know y in advance. Gács and Körner gave an example where mutual information cannot be presented as common information (a third message easily extractable from both x and y). Then this question was studied in the framework of algorithmic information theory by An. Muchnik and A. Romashchenko who found many other examples of this type. K. Makarychev and Yu. Makarychev found a new proof of Gács–Körner results by means of conditionally independent random variables. The question about the difference between mutual and common information can be studied quantitatively: for a given x and y we look for three messages a , b , c such that a and c are enough to reconstruct x , while b and c are enough to reconstruct y .

In this paper:

- We state and prove (using hypercontractivity of product spaces) a quantitative version of Gács–Körner theorem;
- We study the tradeoff between $|a|$, $|b|$, $|c|$ for a random pair (x, y) such that Hamming distance between x and y is εn (our bounds are almost tight);
- We construct “the worst possible” distribution on (x, y) in terms of the tradeoff between $|a|$, $|b|$, $|c|$.

1 Introduction

Let us start with a specific communication problem. Say we have two n -bit strings x and y such that Hamming distance between them is exactly εn . We want to deliver x and y to the corresponding network nodes (Fig. 1). For that we send a common message c to both nodes,

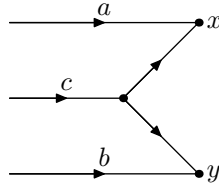


Fig. 1. Network information transmission problem

and separate messages a and b to each node. What tradeoff between the lengths of a , b and c is possible (if we are willing to be able to transmit any such pair (x, y))? If we require c to be empty, then, clearly, $|a|, |b| \geq n$. On the other hand, if a and b are empty, then $|c| \geq (1 + H(\varepsilon) + o(1))n$. In this paper we study this tradeoff and obtain almost tight bounds for $|a|$, $|b|$ and $|c|$.

We can consider more general situation. Suppose we have two dependent random variables X and Y . We sample n independent copies of (X, Y) (let us call them (X^1, Y^1) , (X^2, Y^2) , \dots , (X^n, Y^n)) and want to transmit $x := X^1 X^2 \dots X^n$ and $y := Y^1 Y^2 \dots Y^n$. We want our

* Supported in part by NAFIT ANR-08-EMER-008-01, RFBR 09-01-00709, and RFBR 10-01-93109 grants

transmission to be successful with probability close to 1. Again we are interested in the tradeoff between $|a|$, $|b|$ and $|c|$. Information Theory gives three trivial bounds:

$$|a| + |c| \geq (H(X) + o(1))n, \quad (1)$$

$$|b| + |c| \geq (H(Y) + o(1))n, \quad (2)$$

$$|a| + |b| + |c| \geq (H(X, Y) + o(1))n. \quad (3)$$

This transmission problem was first considered in [1]. In the paper [2] Gács and Körner showed that the first two bounds could not be tight simultaneously (that is, “common information is less than mutual information”) unless the joint distribution of X and Y is in some sense degenerate. Konstantin Makarychev and Yury Makarychev presented in [3] another proof of the same theorem using a notion of conditional independence of random variables.

One can also pose a similar problem in the framework of Algorithmic Information Theory. Suppose we have two strings $x, y \in \{0, 1\}^n$ such that $C(x) \approx n$, $C(y) \approx n$ and $I(x : y) \approx 0.5n$ (here $C(\cdot)$ stands for plain Kolmogorov complexity and $I(\cdot : \cdot)$ — for algorithmic mutual information). Does there exist a string z such that

- $C(z) \leq 0.51n$,
- $C(x | z) \leq 0.51n$,
- $C(y | z) \leq 0.51n$?

Again, it turns out that the answer is negative in general (see [4], [5], [6], [7]). Information-theoretic and Kolmogorov-complexity-theoretic formulations are similar: a stands for the shortest program, which converts z to x , b stands for the shortest program, which converts z to y , and c stands for the shortest program, which outputs z .

All these problems can be reformulated using a simple unifying combinatorial notion of the profile of a bipartite graph. Suppose we have a bipartite graph $E \subseteq X \times Y$. Consider the following communication problem: transmit endpoints of an edge of E .

Definition 1. *We say that a triple (α, β, γ) belongs to the profile of E if there exist two mappings $f: \{0, 1\}^\alpha \times \{0, 1\}^\gamma \rightarrow X$, $g: \{0, 1\}^\beta \times \{0, 1\}^\gamma \rightarrow Y$ such that for every $(x, y) \in E$ there exist $a \in \{0, 1\}^\alpha$, $b \in \{0, 1\}^\beta$, $c \in \{0, 1\}^\gamma$ such that $x = f(a, c)$ and $y = g(b, c)$.*

There is an obvious one-to-one correspondence between the profile of E and the set of possible lengths of three messages that we may transmit if we are willing to be able to deliver endpoints of any edge in E . We can consider profiles not only for graphs, but also for distributions (we require the existence of a, b, c with high probability). If we consider the adjacency matrix M of E , then the abovementioned question can be reformulated as follows: for which triples (α, β, γ) it is possible to cover all ones in M with 2^γ combinatorial rectangles of size $2^\alpha \times 2^\beta$?

In this paper we study profiles of various graphs and distributions. In Section 2 we relate the profile of E and the maximum number of ones in the rectangles in M of a given size. In one direction the relation is obvious, but surprisingly it turns out that in some cases (namely, for edge-transitive graphs) the existence of a rectangle with many ones implies the existence of a good rectangle cover. We also prove several simple bounds on the profile of a bipartite graph which will be useful later. In Section 3 we devise a pretty general tool to upper-bound number of ones in rectangles in M of a given size. It relies on hypercontractivity in product spaces (in [8] hypercontractivity was used for a similar information problem). Using it we prove a quantitative version of Gács–Körner Theorem (for the case of uniform marginal distributions) and improve upon [2] and [3]. In Sections 4 and 5 we prove almost tight bounds for the profile of our first example (strings with Hamming distance εn) using all these tools. Then in Section 6 we build an example of a graph with a minimal possible profile (unsurprisingly, we just prove that random graphs do the job). This example was announced in [7] without proof.

2 Profile and the maximal number of edges in a rectangle

In this section we make two observations. Let $E \subseteq X \times Y$ be a set of edges of a bipartite graph. For given α and β consider combinatorial rectangles $X' \times Y'$ where $X' \subseteq X$ has cardinality 2^α and $Y' \subseteq Y$ has cardinality 2^β . Let $R_E(\alpha, \beta)$ be the maximum number of edges covered by rectangle of this type.

Proposition 1. *If $R_E(\alpha, \beta) \cdot 2^\gamma < |E|$, then the triple (α, β, γ) does not belong to the profile of $|E|$.*

This observation is obvious: each of 2^γ rectangles covers at most $R_E(\alpha, \beta)$ edges, so they cannot cover the entire graph.

It turns out that this observation can be reversed for the case of edge-transitive graphs. Recall that a bipartite graph $E \subseteq X \times Y$ is *edge-transitive* if the group of its automorphisms acts transitively on E (for every two edges e and e' there is an automorphism that maps e to e' ; automorphisms are pairs of permutations $X \rightarrow X$ and $Y \rightarrow Y$ that generate a permutation of E).

Proposition 2. *Assume that E is edge-transitive. If $R_E(\alpha, \beta) \cdot 2^\gamma \geq |E|$, then the triple $(\alpha, \beta, \gamma + \log \log |E|)$ belongs to the profile.*

Proof. Let $R = X' \times Y'$ be a rectangle that contains many edges: let K be the number of edges in it, so that $K \cdot 2^\gamma \geq |E|$. We will cover E by shifted copies of R . Consider m independent randomly chosen automorphisms (all elements of the group are equiprobable); let R_1, \dots, R_m be the images of R under these automorphisms.

We want to show that R_1, \dots, R_m cover E with positive probability. Indeed, for a given edge e the probability of being covered by one R_i is $K/|E|$ (the preimage of e under the automorphism is uniformly distributed in E due to edge transitivity). The probability of *not* being covered is therefore $(1 - K/|E|)$. Different automorphisms are independent, so the probability of e to avoid all R_i is $(1 - K/|E|)^m$. The probability that *some* edge is not covered is bounded by $|E| \cdot (1 - K/|E|)^m$. The assumption guarantees that $K/|E| < 2^{-\gamma}$, so $m = 2^\gamma \log |E|$ is enough to make this probability less than 1. (Indeed, $(1 - 2^{-\gamma})^{2^\gamma} \approx 1/e < 1/2$.)

In the typical application the values of α, β, γ are of the same order of magnitude as $\log |E|$, so $\log \log |E|$ is small compared to α, β, γ .

Now we state and prove several (trivial) bounds on the profile of E .

Proposition 3. *Let $E \subseteq X \times Y$ be a bipartite graph without isolated nodes. Let $0 \leq \alpha \leq \log |X|$ and $0 \leq \beta \leq \log |Y|$.*

1. *If $\alpha + \gamma < \log |X|$ or $\beta + \gamma < \log |Y|$, then $\langle \alpha, \beta, \gamma \rangle$ does not belong to the profile of E .*
2. *If $\alpha + \beta + \gamma < \log |E|$, then $\langle \alpha, \beta, \gamma \rangle$ does not belong to the profile of E .*
3. *If $\min(\alpha, \beta) + \gamma \geq \log |E|$, then $\langle \alpha, \beta, \gamma \rangle$ belongs to the profile of E .*
4. *If $\alpha + \beta + \gamma \geq \log |X| + \log |Y|$, then $\langle \alpha, \beta, \gamma \rangle$ belongs to the profile of E .*

Proof. 1. If $\alpha + \gamma < \log |X|$, then we are unable to cover X (here we use that there are no isolated nodes in E). The second case is similar.

2. If $\alpha + \beta + \gamma < \log |E|$, then it is obviously impossible to cover all edges.

3. Using one $2^\alpha \times 2^\beta$ rectangle we can cover *any* $2^{\min(\alpha, \beta)}$ 1's. Thus, we can cover E with $|E|/2^{\min(\alpha, \beta)}$ rectangles.

4. If $\alpha + \beta + \gamma \geq \log |X| + \log |Y|$, then we can cover not only E , but the entire $X \times Y$.

3 Upper-bounding $R_E(\alpha, \beta)$

To apply Proposition 1 to regular graphs, we need a technique to upper-bound $R_E(\alpha, \beta)$. Let us consider slightly more general situation. Instead of a regular graph $E \subseteq X \times Y$ let us consider a distribution \mathcal{D} over $X \times Y$ such that its marginal distributions are uniform. Then the natural generalization of $R_E(\alpha, \beta)/|E|$ is the following quantity:

$$R_{\mathcal{D}}(\alpha, \beta) := \max_{\substack{|A| \leq 2^\alpha \\ |B| \leq 2^\beta}} \Pr_{(x,y) \sim \mathcal{D}} [x \in A, y \in B].$$

Now let us generalize our problem even more. Suppose that \mathcal{D} 's marginal distributions are not necessarily uniform. Let us denote them by \mathcal{D}_X and \mathcal{D}_Y . Suppose we have two sets $A \subseteq X$, $B \subseteq Y$ such that $\Pr_{x \sim \mathcal{D}_X} [x \in A] \leq \mu$, $\Pr_{y \sim \mathcal{D}_Y} [y \in B] \leq \nu$. We are interested in upper bounds on $\Pr_{(x,y) \sim \mathcal{D}} [x \in A, y \in B]$. There is an obvious bound $\min\{\mu, \nu\}$ for this quantity, but as we will see if \mathcal{D} is in some sense non-degenerate, then we can sharpen this bound. From now on we assume that $\text{supp } \mathcal{D}_X = X$ and $\text{supp } \mathcal{D}_Y = Y$ (otherwise, we can reduce either X or Y).

Let us call \mathcal{D} *non-degenerate* if $\text{supp } \mathcal{D}$ is a connected bipartite graph on (X, Y) and *degenerate* otherwise.

To formulate the upper bound, we shall introduce a parameter $\delta(\mathcal{D})$ with the following properties:

- $\delta(\mathcal{D}) > 0$ iff \mathcal{D} is non-degenerate (Theorem 1 below);
- $\delta(\mathcal{D}_1 \otimes \mathcal{D}_2) \geq \min\{\delta(\mathcal{D}_1), \delta(\mathcal{D}_2)\}$ (Theorem 2 below), here $\mathcal{D}_1 \otimes \mathcal{D}_2$ is a product distribution of \mathcal{D}_1 and \mathcal{D}_2 ;
- if $\delta(\mathcal{D}) > 0$, then there is a non-trivial upper bound on $\Pr_{(x,y) \sim \mathcal{D}} [x \in A, y \in B]$ (Theorem 3 below).

If \mathcal{D} 's matrix was symmetric, then one could in principle define $\delta(\mathcal{D}) := 1 - \lambda(\mathcal{D})$, where $\lambda(\mathcal{D})$ is the second largest eigenvalue of \mathcal{D} 's matrix (the largest eigenvalue is clearly 1). Then, obviously, all three desired properties are true (the third one is a corollary of Expander Mixing Lemma [9]). The problem is that Expander Mixing Lemma is too weak for our purposes (especially if $\mu, \nu = o(1)$), so we need something stronger.

Now let us define $\delta(\mathcal{D})$. Consider $\mathcal{F}_X, \mathcal{F}_Y$ — linear spaces of \mathbb{R} -valued functions on X and Y respectively. Consider the following linear operator $T_{\mathcal{D}}: \mathcal{F}_Y \rightarrow \mathcal{F}_X$:

$$(T_{\mathcal{D}}f)(x) := \mathbb{E}_{(x',y) \sim \mathcal{D}} [f(y) \mid x = x'].$$

Let us consider standard L_p -norms on \mathcal{F}_X and \mathcal{F}_Y : $\|f\|_p := \mathbb{E}[|f|^p]^{1/p}$ (here expectation is taken over X or Y with respect to $\mathcal{D}_X, \mathcal{D}_Y$ respectively). It is easy to check that $T_{\mathcal{D}}$ is an L_p -contraction for all $1 \leq p \leq \infty$.

Lemma 1. $\|T_{\mathcal{D}}f\|_p \leq \|f\|_p$ for all $1 \leq p \leq \infty$ and $f \in \mathcal{F}_Y$.

$\delta(\mathcal{D})$ characterizes to what extent Lemma 1 can be sharpened for a particular \mathcal{D} near $p = 2$.

Definition 2.

$$\delta(\mathcal{D}) := \max \left\{ \delta \leq 1 \mid \forall f \in \mathcal{F}_Y \quad \|T_{\mathcal{D}}f\|_{2+\frac{\delta}{1-\delta}} \leq \|f\|_{2-\delta} \right\}$$

This definition makes sense because the following Lemma holds.

Lemma 2. *If $1 \leq p \leq q \leq \infty$, then*

$$\|f\|_p \leq \|f\|_q,$$

there is an equality iff f is constant.

Now let us restate the properties of $\delta(\mathcal{D})$.

Theorem 1. $\delta(\mathcal{D}) > 0$ iff \mathcal{D} is non-degenerate.

Theorem 2. $\delta(\mathcal{D}_1 \otimes \mathcal{D}_2) \geq \min\{\delta(\mathcal{D}_1), \delta(\mathcal{D}_2)\}$.

Theorem 3. *Let $A \subseteq X$, $B \subseteq Y$. If $\Pr_{x \sim \mathcal{D}_X}[x \in A] \leq \mu$ and $\Pr_{y \sim \mathcal{D}_Y}[y \in B] \leq \nu$, then*

$$\Pr_{(x,y) \sim \mathcal{D}}[x \in A, y \in B] \leq \mu\nu + \left(\mu^{2-\delta(\mathcal{D})}(1-\mu) + \mu(1-\mu)^{2-\delta(\mathcal{D})}\right)^{1/(2-\delta(\mathcal{D}))} \cdot \left(\nu^{2-\delta(\mathcal{D})}(1-\nu) + \nu(1-\nu)^{2-\delta(\mathcal{D})}\right)^{1/(2-\delta(\mathcal{D}))}.$$

For the proof of Theorem 2 see [10]. Theorem 1 and Theorem 3 will be proved in the Appendix. We will typically apply Theorem 3 in situations, where μ and ν tend to zero. Let us instantiate Theorem 3 for this case.

Corollary 1. *If $\mu = o(1)$ and $\nu = o(1)$ then*

$$\Pr_{(x,y) \sim \mathcal{D}}[x \in A, y \in B] \leq O\left((\mu\nu)^{1/(2-\delta(\mathcal{D}))}\right).$$

Now we state and prove a quantitative version of Gács–Körner theorem [2] for the case of uniform marginal distributions.

Theorem 4. *Let \mathcal{D} be a distribution over $X \times Y$ with uniform marginal distributions. We sample n independent copies $(X^1, Y^1), (X^2, Y^2), \dots, (X^n, Y^n)$ of \mathcal{D} and want to transmit $x := X^1 X^2 \dots X^n$ and $y := Y^1 Y^2 \dots Y^n$ as in Fig. 1 (with probability $1 - 2^{-\Omega(n)}$). Then,*

$$|a| + |b| + (2 - \delta(\mathcal{D}))|c| \geq (\log |X| + \log |Y| + o(1))n.$$

That is, if \mathcal{D} is non-degenerate, then the bounds (1) and (2) could not be tight simultaneously, since $\delta(\mathcal{D}) > 0$, $H(X) = \log |X|$, and $H(Y) = \log |Y|$.

The proof is a trivial combination of Theorem 2, Corollary 1, and Proposition 1.

4 Fixed distance graph and its rectangles

In this section we use Propositions 1 and 2 and the results of Section 3 to analyze the transmission of two strings with Hamming distance εn (see the beginning of Section 1).

Let us consider the following bipartite graph $G_{n,\varepsilon} \subseteq \{0, 1\}^n \times \{0, 1\}^n$. There is an edge (x, y) iff Hamming distance between x and y is exactly εn .

We are interested in the profile of $G_{n,\varepsilon}$. Let us denote $M_{n,\varepsilon}$ its adjacency matrix. For simplicity we restrict ourselves to the case where α and β (lengths of messages a and b sent to each node separately) are equal: $\alpha = \beta = \tau n$, $\gamma = \varkappa n$, where $0 < \tau < 1$, $0 < \varkappa < 1 + H(\varepsilon)$ are constants. Let us denote by $\Lambda(\varepsilon, \tau)$ the following fraction:

$$\Lambda(\varepsilon, \tau) = \lim_{n \rightarrow \infty} \frac{\log(R(\varepsilon, \tau, n) / |G_{n,\varepsilon}|)}{n},$$

where $R(\varepsilon, \tau, n)$ is the maximum number of 1's in a rectangle of $M_{n,\varepsilon}$ of size $2^{\tau n} \times 2^{\tau n}$, and $|G_{n,\varepsilon}|$ is the total number of edges in $G_{n,\varepsilon}$. Since $G_{n,\varepsilon}$ is edge-transitive, bounds for Λ can be directly translated into profile bounds (via Propositions 1 and 2).

Let us state our main combinatorial result:

Theorem 5. (*Lower bound*) If $\tau < H(1 - \sqrt{1 - \varepsilon})$, then $\Lambda(\varepsilon, \tau) \geq -(1 + H(\varepsilon) - 2\tau)$. If $\tau \geq H(1 - \sqrt{1 - \varepsilon})$, then

$$\Lambda(\varepsilon, \tau) \geq -\left(1 + H(\varepsilon) - \tau - \alpha H\left(\frac{\varepsilon}{2\alpha}\right) - (1 - \alpha)H\left(\frac{\varepsilon}{2(1 - \alpha)}\right)\right),$$

where $0 < \alpha < 1/2$ and $H(\alpha) = \tau$.

(*Upper bound*) $\Lambda(\varepsilon, \tau) \leq -\frac{1 - \tau}{1 - \varepsilon}$.

The lower bound on $\Lambda(\varepsilon, \tau)$

To prove that $\Lambda(\varepsilon, \tau)$ is large we show that some rectangle in $M_{n,\varepsilon}$ has many 1's. Indeed, consider a rectangle $C \times C$, where elements of C are strings that contain exactly αn ones; obviously there are $2^{(H(\alpha) + o(1))n}$ of them (i.e. $\tau = H(\alpha) + o(1)$). The number of 1's in $C \times C$ equals

$$\binom{n}{\alpha n} \binom{\alpha n}{\varepsilon n/2} \binom{(1 - \alpha)n}{\varepsilon n/2} = 2^{(H(\alpha) + \alpha H(\varepsilon/2\alpha) + (1 - \alpha)H(\varepsilon/2(1 - \alpha)) + o(1))n}.$$

The total number of 1's in $M_{n,\varepsilon}$ is

$$2^n \binom{n}{\varepsilon n} = 2^{(1 + H(\varepsilon) + o(1))n}.$$

Thus

$$\begin{aligned} \Lambda(\varepsilon, \tau) &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \log \frac{2^{(H(\alpha) + \alpha H(\varepsilon/2\alpha) + (1 - \alpha)H(\varepsilon/2(1 - \alpha)) + o(1))n}}{2^{(1 + H(\varepsilon) + o(1))n}} = \\ &= -(1 + H(\varepsilon) - H(\alpha) - \alpha H(\varepsilon/2\alpha) - (1 - \alpha)H(\varepsilon/2(1 - \alpha))). \end{aligned}$$

If $\tau = H(1 - \sqrt{1 - \varepsilon})$, then the sphere gives a rectangle with $2^{(2\tau + o(1))n}$ ones, which is clearly optimal. For $\tau' < \tau$ we can subsample this rectangle and get $2^{\tau'n} \times 2^{\tau'n}$ rectangle with $2^{(2\tau' + o(1))n}$ ones.

The upper bound on $\Lambda(\varepsilon, \tau)$

Let us prove that $\Lambda(\varepsilon, \tau) \leq -\frac{1 - \tau}{1 - \varepsilon}$. Consider the distribution $D_{n,\varepsilon}$ on $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$: x is uniformly distributed in $\{0, 1\}^n$; and $y \in \{0, 1\}^n$ is obtained from x by independently changing each bit with probability ε .

This distribution generates an edge in $G_{n,\varepsilon}$ with probability at least $1/n$, and all the edges of $G_{n,\varepsilon}$ are equiprobable. So instead of counting the number of edges in a rectangle, we may estimate the $D_{n,\varepsilon}$ -probability of this rectangle (the factor n does not matter with our precision). It is enough to show, therefore, that for every $C_1, C_2 \subseteq \{0, 1\}^n$ such that $|C_1| = |C_2| = 2^{\tau n}$ the following inequality holds:

$$\Pr_{(x,y) \sim D_{n,\varepsilon}} [(x, y) \in C_1 \times C_2] \leq 2^{-(\frac{1 - \tau}{1 - \varepsilon} + o(1))n}. \quad (4)$$

If we show that $\delta(D_{n,\varepsilon}) \geq 2\varepsilon$, then we can plug this bound into Corollary 1, and obtain (4). Since $D_{n,\varepsilon} = D_{1,\varepsilon}^{\otimes n}$, using Theorem 2 one can reduce this statement to the following well-known inequality.

Theorem 6 (Two-Point Inequality).

$$\delta(D_{1,\varepsilon}) \geq 2\varepsilon.$$

For the proof see [10].

5 The profile of the fixed-distance graph

First, we use Theorem 5 and Propositions 2 and 1 to get explicit bounds for the combinatorial profile of the fixed-distance graph $G_{n,\varepsilon}$ that improve those given by Proposition 3.

Theorem 7. *Let $0 < \tau, \varkappa < 1$ be constants.*

- *If $\varkappa < (1 - \tau)/(1 - \varepsilon)$, then for sufficiently large n the triple $\langle \tau n, \tau n, \varkappa n \rangle$ does not belong to the profile of $G_{n,\varepsilon}$.*
- *There are two following “positive bounds”.*
 - *If $\tau < H(1 - \sqrt{1 - \varepsilon})$ and $\varkappa > 1 + H(\varepsilon) - 2\tau$, then for sufficiently large n the triple $\langle \tau n, \tau n, \varkappa n \rangle$ belongs to the profile of $G_{n,\varepsilon}$.*
 - *If $\tau \geq H(1 - \sqrt{1 - \varepsilon})$ and*

$$\varkappa > 1 + H(\varepsilon) - \tau - \alpha H\left(\frac{\varepsilon}{2\alpha}\right) - (1 - \alpha)H\left(\frac{\varepsilon}{2(1 - \alpha)}\right),$$

where $0 < \alpha < 1/2$ and $H(\alpha) = \tau$, then for sufficiently large n the triple $\langle \tau n, \tau n, \varkappa n \rangle$ belongs to the profile of $G_{n,\varepsilon}$.

Figure 2 shows the bounds for the profile for $\varepsilon = 0.11\dots$ (for this value the total number of edges is $2^{1.5n}$). It shows trivial upper and lower bounds from Proposition 3, as well as our results (Theorem 5).

Not that our bounds are tight in two regions:

- if $\tau < H(1 - \sqrt{1 - \varepsilon})$, then our upper bound is equal to the trivial lower bound;
- if $\tau = 1 - o(1)$, then our upper bound is asymptotically equal to our lower bound.

The same bounds can be obtained for other notions of profile, so the results are directly comparable with previous work. Let us show how this can be done for Kolmogorov complexity.

Let us assume that for every n a bipartite graph $E_n \subseteq \{0, 1\}^n \times \{0, 1\}^n$ is fixed (and there is an algorithm computing E_n given n). Let α, β, γ be some positive rational numbers. Let $R_n(\alpha, \beta)$ be the maximum number of edges covered by a rectangle $X' \times Y'$ with $|X'| = 2^{\alpha n}$ and $|Y'| = 2^{\beta n}$.

Proposition 4. *There exists some constant d such that:*

- *If $R_n(\alpha, \beta) \cdot 2^{\gamma n} \leq |E_n|$, then for all sufficiently large n for most edges $(x, y) \in E_n$ there is no string c such that*

$$C(c) < \gamma n - d \log n, \quad C(x|c) < \alpha n - d \log n, \quad C(y|c) < \beta n - d \log n.$$

- *If E_n is edge-transitive and $R_n(\alpha, \beta) \cdot 2^{\gamma n} \geq |E_n|$, then for all sufficiently large n for every edge $(x, y) \in E_n$ there exists a string c such that*

$$C(c) < \gamma n + d \log n, \quad C(x|c) < \alpha n + d \log n, \quad C(y|c) < \beta n + d \log n.$$

Proof. For each c we consider a rectangle $X' \times Y'$ where X' consists of strings x such that $C(x|c) < \alpha n$ and Y' consists of strings y such that $C(y|c) < \beta n$ ($|X'| < 2^{\alpha n}$, $|Y'| < 2^{\beta n}$). The number of strings c such that $C(c) < \gamma n - d \log n$ is less than $2^{\gamma n}/n^d$, so the edges covered by these rectangles form a minority ($1/n^d$ -fraction); for all other pairs there is no string c with described properties.

The second part of the theorem is also easy. Proposition 2 guarantees that for sufficiently large n the set $|E_n|$ can be covered by $2^{\gamma n + O(\log n)}$ rectangles of size $2^{\alpha n} \times 2^{\beta n}$. By exhaustive

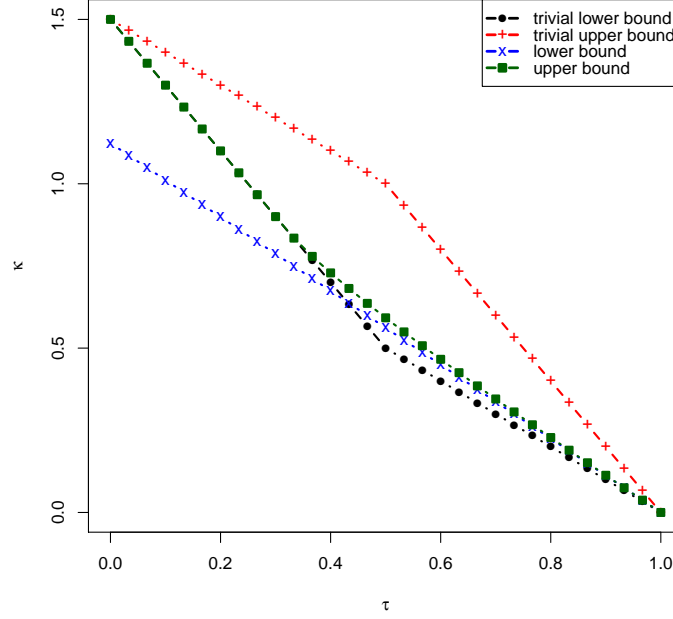


Fig. 2. bounds on the profile from Proposition 3 and Theorem 5, $d(x, y) = \varepsilon n$, $\varepsilon = 0.11 \dots$

search, we can find first covering of this sort (in some natural order), and this covering is determined by n , so its complexity is $O(\log n)$. Then we let c be the number of the rectangle $X' \times Y'$ that covers a given edge (x, y) . The complexity of c is at most $\gamma n + O(\log n)$. Knowing c (and the entire covering, which has complexity $O(\log n)$) we can describe x and y by their numbers in X' and Y' .

One can also show that a random pair (x, y) generated with distribution $D_{n, \varepsilon}$ will have a profile (in terms of complexity) within the bounds from Theorem 4 (with high probability). Indeed, the law of large number says that the number of places where x and y differ is close to εn , and for each fixed number of difference we get a uniformly random edge in $G_{n, \varepsilon'}$ for ε' close to ε . It remains to note that our bounds are continuous (as functions of ε).

Let us note for comparison that (a weaker) upper bound for this distribution can be obtained using conditional independence technique from [6]. It is quite involved, but it also has a form $\kappa \geq c(\varepsilon)(1 - \tau)$ (where $c(\cdot)$ is an explicit (but cumbersome) function) as in Theorem 7. So, let us compare $c(\varepsilon)$ with $1/(1 - \varepsilon)$ (the bigger value is better) for different values of ε (see Fig. 3).

ε	$1/(1 - \varepsilon)$	$c(\varepsilon)$
0.1	1.11...	1.000015...
0.2	1.25	1.016...
0.3	1.43...	1.067...
0.4	1.67...	1.33...

Fig. 3. A comparison of upper bounds from Theorem 5 and from [6]

6 A stochastic pair with minimal profile

A pair with minimal profile is constructed in [7] using the technique developed by An. Muchnik [4], [5]. However, this construction is quite artificial and cannot be translated into Shannon information theory, because the constructed pair is not a typical object in a simple family (is not stochastic in the sense of algorithmic information theory). In this section we show how to construct a graph with minimal combinatorial profile (for graphs with given number of edges). For that we first prove (using probabilistic arguments) that such a graph exists; after that it can be found by brute-force search. This gives us a stochastic pair with minimal profile. This result was announced in [7] as an unpublished result of An. Muchnik (1958–2007) and was not published since then.

To analyze random graphs we need a version of Chernoff inequality which deals with negatively correlated random variables [11].

Theorem 8 (Chernoff inequality). *Let X_1, X_2, \dots, X_n be negatively correlated binary random variables (i.e., for every $i_1 < i_2 < \dots < i_k$ we have $\Pr[X_{i_1} = 1 \wedge X_{i_2} = 1 \wedge \dots \wedge X_{i_k} = 1] \leq \prod_{j=1}^k \Pr[X_{i_j} = 1]$).*

Let $\mu = \mathbb{E}[X_1 + X_2 + \dots + X_n]$ and $\delta > 0$. Then

$$\Pr[X_1 + X_2 + \dots + X_n \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu.$$

Following [7], we consider (for simplicity) graphs with 2^n vertices in each part and $2^{1.5n}$ edges. We choose a random graph of this type (i.e., uniformly choose a matrix of size $2^n \times 2^n$ with $2^{1.5n}$ ones).

Theorem 9. *If $\kappa + \tau < 1.5$ and $\kappa + 2\tau < 2$, then for some $\varepsilon > 0$ the probability of the event “every $2^{\tau n} \times 2^{\tau n}$ rectangle in M contains less than $2^{(1.5-\kappa-\varepsilon)n}$ ones” is close to 1.*

Proof. Let us choose $0 < \varepsilon < \min(2 - \kappa - 2\tau, 1.5 - \kappa - \tau)$. The entries of M are negatively correlated, so the Chernoff inequality (with $1 + \delta = 2^{(2-\kappa-2\tau-\varepsilon)n}$ and $\mu = 2^{(2\tau-0.5)n}$) guarantees that for a fixed rectangle of size $2^{\tau n} \times 2^{\tau n}$ the probability of the event “this rectangle contains more than $2^{(1.5-\kappa-\varepsilon)n}$ ones” is bounded by

$$\left(\frac{e^{2^{(2-2\tau-\kappa-\varepsilon)n}-1}}}{2^{(2-2\tau-\kappa-\varepsilon)n} 2^{(2-2\tau-\kappa-\varepsilon)n}} \right)^{2^{(2\tau-0.5)n}} = 2^{-2^{(1.5-\kappa-\varepsilon+o(1))n}}$$

Therefore the probability that some $2^{\tau n} \times 2^{\tau n}$ rectangle has too many ones does not exceed

$$\left(\frac{2^n}{2^{\tau n}} \right)^2 \cdot 2^{-2^{(1.5-\kappa-\varepsilon+o(1))n}} \leq 2^{2^{(\tau+o(1))n}} \cdot 2^{-(1.5-\kappa-\varepsilon+o(1))n}.$$

Since $\varepsilon < 1.5 - \kappa - \tau$, we are done.

Remark. Note that it follows from Propositions 1 and 3 that the bounds in Theorem 9 are the best possible.

Another simple observation: with high probability all rows and columns of M contain at most $2 \cdot 2^{0.5n}$ elements. Indeed, the probability that in a given row (or column) the number of ones is twice more than the expected value $2^{0.5n}$, is doubly exponentially small (it follows from theorem 8), and we have only exponentially many rows and columns.

Then we can follow the plan described above: we conclude that there is a graph with both properties and logarithmic complexity; then, using the first part of proposition 4 one can easily see that a typical edge (x, y) of this graph has minimal profile in the sense explained in [7]. (The second property is needed to show that the complexities of x and y in a typical edge are close to n .)

7 Acknowledgements

I would like to thank Andrei Romashchenko and Alexander Shen for posing the problem and for fruitful discussions, and Alex Samorodnitsky for useful advice.

References

1. Robert Gray and Aaron Wyner. Source coding over simple networks. *Bell Systems Technical Journal*, 53(9):1681–1721, 1974.
2. Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.
3. Konstantin Makarychev and Yuri Makarychev. Conditionally independent random variables. Preprint, 2005.
4. Andrej Muchnik. On the extraction of common information of two words. *Pervyi vsemirnyi kongress obshchestva matematicheskoi statistiki i teorii veroyatnostei imeni Bernoulli. Tezisy*, 1:453, 1986.
5. Andrej Muchnik. On common information. *Theoretical Computer Science*, 207(2):319–328, 1998.
6. Andrei Romashchenko. Pairs of words with nonmaterializable mutual information. *Problems of Information Transmission*, 36:1–18, 2000.
7. Alexey Chernov, Andrej Muchnik, Andrei Romashchenko, Alexander Shen, and Nikolai Vereshchagin. Upper semi-lattice of binary strings with the relation “ x is simple conditional to y ”. *Theoretical Computer Science*, 271:69–95, 2002.
8. Rudolf Ahlswede and Peter Gács. Spreading of Sets in Product Spaces and Hypercontraction of the Markov Operator. *The Annals of Probability*, 4(6):925–939, 1976.
9. Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the AMS*, 43:439–561, 2006.
10. Christophe Garban and Jeffrey Steif. Lectures on noise sensitivity and percolation. Preprint, 2011.
11. Alessandro Panconesi and Aravind Srinivasan. Randomized Distributed Edge Coloring via an Extension of the Chernoff-Hoeffding Bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.

A Properties of $\delta(\mathcal{D})$

We prove properties of $\delta(\mathcal{D})$ that were promised in Section 3.

Theorem 10. $\delta(\mathcal{D}) > 0$ iff \mathcal{D} is non-degenerate.

Proof. Suppose \mathcal{D} is degenerate. Then there are non-trivial partitions $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$ such that $\Pr_{(x,y) \sim \mathcal{D}}[x \in X_1, y \in Y_2] = 0$ and $\Pr_{(x,y) \sim \mathcal{D}}[x \in X_2, y \in Y_1] = 0$. Consider the indicator $I_{Y_1} \in \mathcal{F}_Y$ of Y_1 that maps Y_1 to 1 and $Y - Y_1$ to 0. Then for every $1 \leq p \leq \infty$ the following equality holds: $\|T_{\mathcal{D}}I_{Y_1}\|_p = \|I_{Y_1}\|_p$. On the other hand by Lemma 2

$$\begin{aligned} \frac{d}{dp} \|I_{Y_1}\|_p &> 0, \\ \frac{d}{dp} \|T_{\mathcal{D}}I_{Y_1}\|_p &> 0, \end{aligned}$$

since $\text{supp } \mathcal{D}_X = X$, $\text{supp } \mathcal{D}_Y = Y$, and I_{Y_1} is non-constant. Thus, clearly, $\delta(\mathcal{D}) = 0$.

Conversely, let \mathcal{D} be non-degenerate. Consider the unit sphere S in $L_2(Y)$. For $f \in S$ let us define $\delta(f) := \max \left\{ \delta \leq 1 \mid \|T_{\mathcal{D}}f\|_{2+\frac{\delta}{1-\delta}} \leq \|f\|_{2-\delta} \right\}$. Since $\delta(\mathcal{D}) = \inf_{f \in S} \delta(f)$, $\delta(f)$ is continuous, and S is compact, it remains to prove that $\delta(f) > 0$ for every $f \in S$. If f is constant then, the inequality is obvious since $\|f\|_p$ is constant. If f is not constant, then the inequality is true, since $\|T_{\mathcal{D}}f\|_2 < \|f\|_2$, so there exists $\eta > 0$ such that $\|T_{\mathcal{D}}f\|_{2+\eta} < \|f\|_{2-\eta}$ ($\|\cdot\|_p$ is continuous in p).

Theorem 11. Let $A \subseteq X$, $B \subseteq Y$. If $\Pr_{x \sim \mathcal{D}_X} [x \in A] = \mu$ and $\Pr_{y \sim \mathcal{D}_Y} [y \in B] = \nu$, then

$$\begin{aligned} \Pr_{(x,y) \sim \mathcal{D}} [x \in A, y \in B] &\leq \mu\nu + \\ &+ \left(\mu^{2-\delta(\mathcal{D})}(1-\mu) + \mu(1-\mu)^{2-\delta(\mathcal{D})} \right)^{1/(2-\delta(\mathcal{D}))} \cdot \left(\nu^{2-\delta(\mathcal{D})}(1-\nu) + \nu(1-\nu)^{2-\delta(\mathcal{D})} \right)^{1/(2-\delta(\mathcal{D}))}. \end{aligned} \quad (5)$$

Proof. Let I_A and I_B be indicators of A and B respectively.

$$\Pr_{(x,y) \sim \mathcal{D}} [x \in A, y \in B] = \mathbb{E}_{(x,y) \sim \mathcal{D}} [I_A(x)I_B(y)]$$

Let us denote $f(x) := I_A(x) - \mu$, $g(x) := I_B(x) - \nu$.

$$\mathbb{E}_{(x,y) \sim \mathcal{D}} [I_A(x)I_B(y)] = \mu\nu + \mathbb{E}_{(x,y) \sim \mathcal{D}} [f(x)g(y)] = \mu\nu + \mathbb{E}_{x \sim \mathcal{D}_X} [f(x)T_{\mathcal{D}}g(x)]$$

By Hölder's inequality

$$\mu\nu + \mathbb{E}_{x \sim \mathcal{D}_X} [f(x)T_{\mathcal{D}}g(x)] \leq \mu\nu + \|T_{\mathcal{D}}f\|_{2+\frac{\delta(\mathcal{D})}{1-\delta(\mathcal{D})}} \|g\|_{2-\delta(\mathcal{D})}.$$

By definition of $\delta(\mathcal{D})$

$$\mu\nu + \|T_{\mathcal{D}}f\|_{2+\frac{\delta(\mathcal{D})}{1-\delta(\mathcal{D})}} \|g\|_{2-\delta(\mathcal{D})} \leq \mu\nu + \|f\|_{2-\delta(\mathcal{D})} \|g\|_{2-\delta(\mathcal{D})}. \quad (6)$$

Let us recall that $f = I_A - \mu$, $g = I_B - \nu$. Thus,

$$\|f\|_{2-\delta(\mathcal{D})} = \left(\mu(1-\mu)^{2-\delta(\mathcal{D})} + \mu^{2-\delta(\mathcal{D})}(1-\mu) \right)^{1/(2-\delta(\mathcal{D}))}, \quad (7)$$

$$\|g\|_{2-\delta(\mathcal{D})} = \left(\nu(1-\nu)^{2-\delta(\mathcal{D})} + \nu^{2-\delta(\mathcal{D})}(1-\nu) \right)^{1/(2-\delta(\mathcal{D}))}. \quad (8)$$

Plugging (7), (8) into (6) we obtain (5).